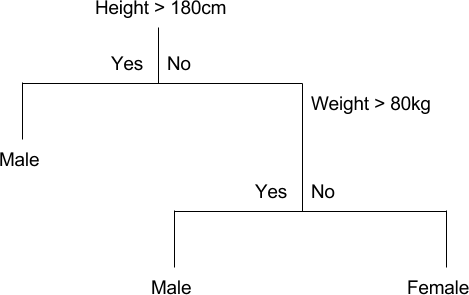
1. Prediction of Categorical Data using Rule base classification and decision tree classification through WEKA using any datasets. Compare the accuracy using two algorithm and plot the graph
2. Create the following dataset using arff file format. To perform cluster analysis using K- Means in WEKA. To change the cluster size and plot the graph and illustrate the visualization of cluster.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EmployeID | Gender | Age | Salary | Credit |
| 111 | Male | 28 | 150000 | 39 |
| 222 | Male | 25 | 150000 | 27 |
| 333 | Female | 26 | 160000 | 42 |
| 444 | Female | 25 | 160000 | 40 |
| 555 | Female | 30 | 170000 | 64 |
| 666 | Male | 29 | 200000 | 72 |

1. Consider this Decision tree :
2. create the data set for the below tree using ARFF format and calculate accuracy and decision for the same
3. Using this decision tree generate the rules based on rule based induction.
4. Compare both the algorithms and plot the confusion matrix.



1. Create an ARFF file for the table below and implement for the Apriori Algorithm and FP growth algorithm and compare the rules generated by both the algorithms. Identify the unique rules generated by the above algorithms.

NOTE: Assume Min\_sup=2 and confidence= 50%

|  |  |
| --- | --- |
| T.ID | ITEMS |
| T1 | SONY, BPL, LG |
| T2 | BPL, SAMSUNG |
| T3 | BPL, ONIDA |
| T4 | SONY, BPL, SAMSUNG |
| T5 | SONY, ONIDA |
| T6 | BPL, ONIDA |
| T7 | SONY, ONIDA |
| T8 | SONY, BPL, ONIDA, LG |
| T9 | SONY, BPL, ONIDA |

1. )Let us consider one example to make the calculation method clear. Assume that the minimum and maximum values for the feature F are $50,000 and $100,000 correspondingly. It needs to range *F* from 0 to 1. In accordance with min-max normalization, *v* = $80,

b) Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000

1. min-max normalization by setting min = 0 and max = 1
2. z-score normalization